

生存時間解析におけるGreenwood式標準誤差推定の性能評価：無増悪生存率の各時点ごとの区間推定

著者	榎本 駿平, 西川 哲夫, 西川 正子
著者(英)	Enomoto Shunpei, Nishikawa Tetsuo, Nishikawa Masako
雑誌名	武蔵野大学数理工学センター紀要
号	5
ページ	1-26
発行年	2020-03-01
URL	http://id.nii.ac.jp/1419/00001265/

生存時間解析における Greenwood 式標準誤差推定の性能評価 — 無増悪生存率の各時点ごとの区間推定 —

Performance evaluation of standard error estimation with Greenwood formula in survival analysis : Pointwise interval estimation of the progression-free survival

榎 本 駿 平¹

Enomoto Shunpei

西 川 哲 夫²

Nishikawa Tetsuo

西 川 正 子³

Nishikawa Masako

概要

生存時間解析において、生存率の時点ごとの標準誤差の推定には、その時点までのイベント発現状況のデータを用いて推定する Greenwood 式標準誤差（以下、GWSE と表す）推定値が使われている。生存時間解析では、イベント発現の時点が区間単位でしかわからない場合（区間打ち切りデータ）があり、その場合、区間の右端を用いる右端代入法が主に用いられている。本研究では、区間打ち切りデータが存在するときの GWSE の推定性能とそれを用いた生存率の区間推定の性能評価を目的として、GWSE の推定精度と信頼区間の被覆確率についてシミュレーションによる検討を行った。右端代入法と共に中点代入法と左端代入法についてもパラメータを変えて評価した結果、代入法ごとに被覆確率に特徴があることが分かった。被験者数が多い場合、ほとんどの場合 GWSE は 10000 回の推定によって得られた無増悪生存率の標準偏差（以下、SD と表す）を正しく推定できていたが、信頼区間の被覆確率については、大部分の時点で名目の信頼水準に満たなかった。しかし、最終検査時点においては左端代入法を用いた場合ほとんどのパラメータの組み合わせで良好な結果が得られることが分かった。

1 導入

1.1 背景

イベント発現までの時間の分布を調査する研究では、ある個体ではイベントが発現しないままで観測が打ち切られることがある。そのようなデータは右側打ち切りデータと呼ばれる。イベント発現までの時間を分析する生存時間解析において、生存関数の推定

¹ 武蔵野大学工学部数理工学科

² 武蔵野大学工学部数理工学科特任教授 / 武蔵野大学数理工学センター員

³ 東京慈恵会医科大学臨床研究支援センター教授

には Kaplan-Meier 法（以下、KM 法と表す）が使われている。また、各時点の生存率の信頼区間の構成には Greenwood 式標準誤差（以下、GWSE と表す）推定値が使われている。KM 法とは、観測されたイベントが発現するまでの時間及び観測打ち切りまでの時間（生存時間のデータ）からある時間以上生存する確率を表す生存関数を推定する手法である。時間を t 、イベントの発現時間を昇順に並べたときの j 番目の時間を $t_{(j)}$ 、各時点の生存率推定値を $\hat{S}(t)$ とし $t_{(j)}$ の直前の生存数を N_j 、 $t_{(j)}$ でのイベント数を d_j とすると $\hat{S}(t)$ は以下のようなになる（たとえば、西川(2019), p.29, p.39[1], など）。

・ KM 法

$$\hat{S}(t) = \begin{cases} 1 & t < t_{(1)} \text{ のとき} \\ \prod_{k=1}^j \left(1 - \frac{d_k}{N_k}\right) & t_{(j)} \leq t \leq t_{(j+1)} \text{ のとき} \end{cases}$$

また、GWSE とは 1 つの集団から得られた生存時間のデータを用いて推定した各時点の生存率 $\hat{S}(t)$ の標準誤差で、推定は式（1）による。

・ GWSE

$$\sqrt{\text{Var}(\hat{S}(t))} = \sqrt{\hat{S}(t)^2 \sum_{i: t_{(i)} \leq t} \frac{d_i}{N_i(N_i - d_i)}} \quad (t_{(j)} \leq t < t_{(j+1)} \text{ のとき}) \cdots (1)$$

ある起点からイベント発現までの時間である生存時間のデータには、例えば、腫瘍増悪のように検査をしないとイベント発現の有無が分からず、いつ発生したかは正確には不明で、イベントの正確な発生時点は経時的な検査で発生「無し」であった最後の検査日（観察打ち切りになった区間の左端）と発生「有り」であった最初の検査日（観察打ち切りになった区間の右端）の間、ということではわからない。このように、イベント発現までの時間がある時点から別なある時点までに発生した、という観察打ち切りになった時間の幅（以降、観察打ち切りになった区間、と呼ぶ）として得られる場合、これを区間打ち切りデータと呼ぶ。

観測データにイベント発現までの正確な時間（例、死亡までの時間）と右側打ち切りデータ（イベント発現した時点がある時点以降としかわからないデータ）の他に、区間打ち切りデータ（例、腫瘍増悪までの時間）が含まれるデータは部分的区間打ち切りデータと呼ばれる。

検査時点が臨床試験計画により定期的に計画されている場合であっても、検査時点には前後数日から数ヶ月の許容幅が設定されている。観察開始時点から遅い時点であるほど許容幅は長く設定されていることが多い。また、患者（被験者・個体）が来院しなかったり予定日と異なる日に来院したりすることも珍しくない。観察打ち切りになった区間の右端の検査時点が同じであっても観察打ち切りになった区間は異なる場合も多い。その結果、イベントの発生時間が区間打ち切りデータになり観察打ち切りになった区間は個体に

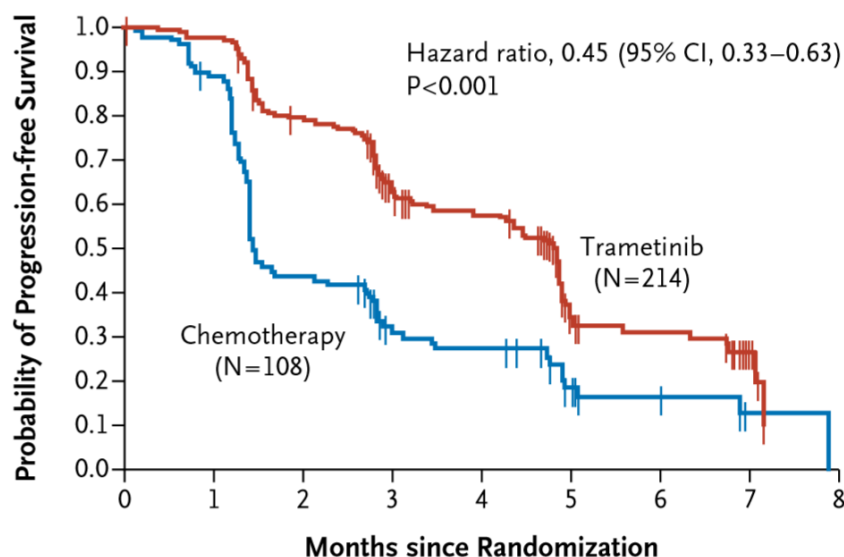
より端点が異なり、長さも異なり、他の個体の観察打ち切りになった区間とのオーバーラップも発生することになる。

区間打ち切りデータに対して、観察打ち切りになった区間内のある 1 点をそのデータの正確な生存時間として扱う 1 点代入法がある。最も簡便な 1 点代入法には、それぞれ区間の左端、中点、右端を代入する左端代入法、中点代入法、右端代入法 (Law and Brookmeyer (1992) [2]) などがある。そのほか、確率的な 1 点代入法 (e. g. Zhang et al. (2009) [3])、期待値代入 (Gauvreau, et al. (1994) [4])、多重代入法 (Rubin (1987) [5]) を利用したアプローチ (Taylor et al. (1990) [6]; Pan (2000) [7] など) などもある。

本稿では以降ほとんど触れないが、区間打ち切りデータに対してある点を代入せず区間打ち切りデータとして解析する方法も研究されており、Sun(2005) [8]には区間打ち切りデータに関する様々なテーマがまとめられている。Ding-Geng Chen et al. (2013) [9]は区間打ち切りデータ解析の応用や解析ソフトについてもまとめられている。日本語での解説は西川(2019)3.4 章[1]、西川(2010, 2018) [10] [11] などがある。

臨床試験への応用では、区間打ち切りデータに対して右端代入法 (腫瘍増悪が認められた検査日当日に増悪が発現したと扱う。実際は、検査日より前に増悪の基準に到達していたかもしれない) を用いてから KM 法により生存関数の推定が行われることが多い。実例として図 1、図 2 に Flaherty ら(2012)[12]の「死亡か増悪のいずれか先に起きたものをイベントとして定義」された無増悪生存率 (集団全体のうち増悪も死亡もしていない集団の割合)、及び Giuliano ら(2017)[13]の「死亡か乳癌の再発のいずれか先に起きたものをイベントとして定義」された無病生存率を示す。

A Progression-free Survival



No. at Risk

Chemotherapy	108	87	43	24	21	10	6	1	0	0
Trametinib	214	205	163	100	88	28	22	5	0	0

図 1 転移性メラノーマ患者の無増悪生存率 (Flaherty et al.(2012)の図 1 より抜粋して引用[12])

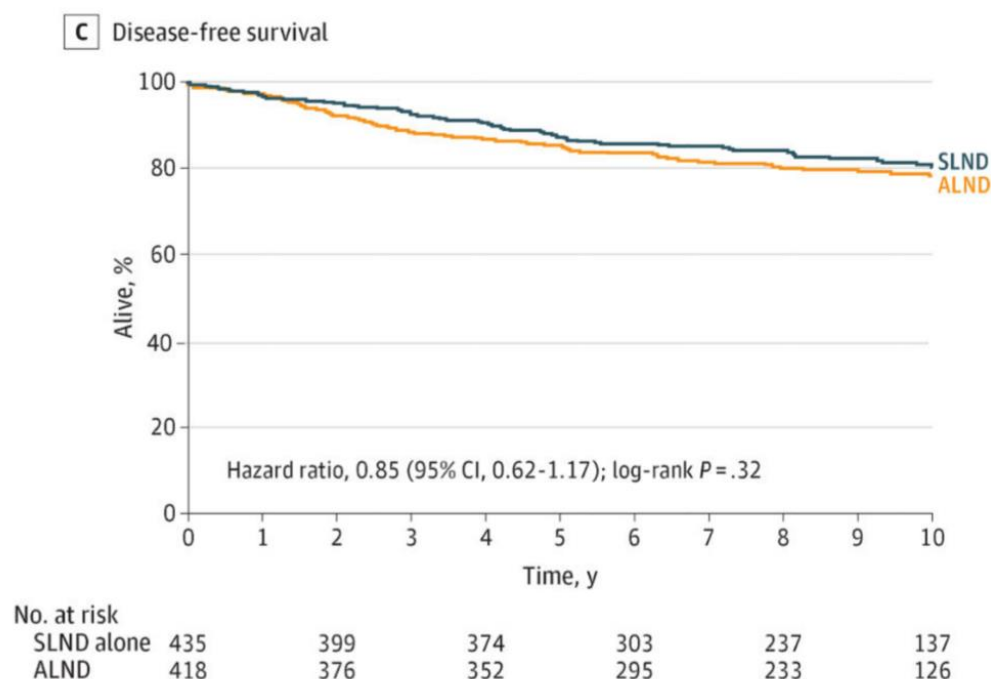


図 2 乳がん患者の無病生存率 (Giuliano et al.(2017)の図 2 より抜粋して引用[13])

1点代入法ではある1点を正確な生存時間として扱うためにGWSEは過小評価になっている。このため、GWSEは各時点の生存率のSDを正しく推定できていない可能性がある。

Law and Brookmeyer (1992)はシミュレーションによる検討を行い、1点代入法では中点代入が良いこと、および観察打ち切りになった区間が広かったり個体によって変化したりする場合などは生存率推定値にバイアスがあることを報告している。

さらに、区間打ち切りデータに対してそれぞれ左端、中点、右端代入法を用いると、個人の無増悪生存時間は左端、中点、右端代入法の順に長くなる。しかし、右側打ち切りデータを含む場合、KM法により無増悪生存率を推定すると、集団としての無増悪生存率は用いる代入法によって左端、中点、右端代入法の順に常になくなるという現象が起こり、直感とは異なる結果が得られることがある。例として、図3にNishikawaら(2003a)[14]の図を示す。実線は中点代入を、破線は右端代入を表している。Nishikawaら(2003b)[15]はこのような現象が起こる十分条件を示した。Nishikawaら(2003a)[14]はこれらを踏まえて臨床試験の設定を模したシミュレーションを行い3種類の1点代入法(左端、中点、右端代入法)によって作成したデータセットをもとにKM法で無増悪生存率推定したときとターンブル法(代入を行わない方法)を、おもに点推定の観点から比較した。右側打ち切りデータが存在するとき、左端、中点、右端それぞれを代入する方法の間では多くの場合には中点代入が平均2乗誤差を小さくすること、および右端代入は必ずしも生存率の過大評価というわけではないことを報告している。

臨床試験の主要な評価時点は観察の最終時点での無増悪生存率等であることが多いので、最終時点の評価に関するシミュレーション研究は散見されるが、近年は、ある時

点までの無増悪生存率曲線下面積 (restricted mean survival time, RMS) も注目されつつある (例えば、Uno et al (2013) [16]、Huang et al (2017) [17])。RMS の点推定には KM 法を、RMS の SE の簡便な推定には観測期間内の時点ごとの GWSE による SE を利用している。

臨床試験において、イベントを増悪または死亡のいずれか先に発現した事象と定義する無増悪生存時間では、例えば図 1 のように、増悪時点は検査時点付近に集中し、時点ごとの無増悪生存率の信頼区間は名目の信頼水準を満たしていないことが予想される。しかし、GWSE の推定性能に関するシミュレーションによる詳細な検討は行われていない。

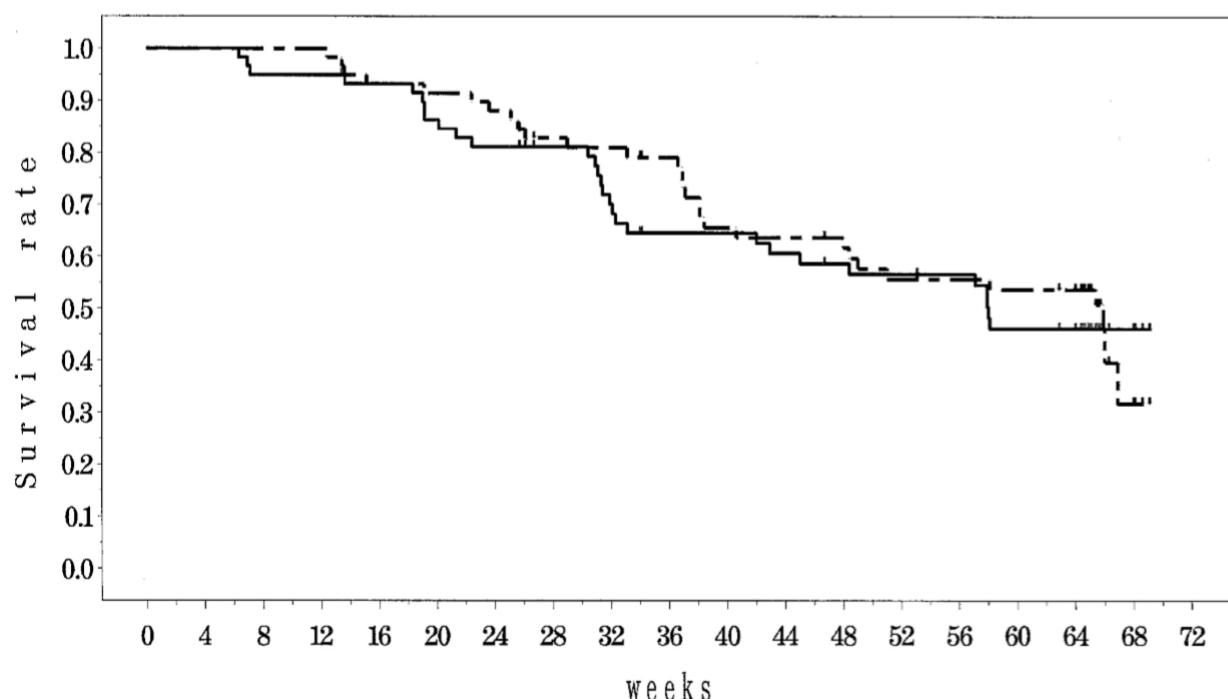


図 3 無増悪生存率 (Nishikawa, M. and Tango, T. (2003a) より図 1 を引用[14])

1.2 目的

本研究では増悪までの時間には 1 点代入法を用いて、無増悪生存時間の生存関数 (無増悪生存率) を KM 法により推定する臨床試験を想定し、3 種類の 1 点代入を行ったデータセットに対し GWSE 推定の性能及び無増悪生存率の区間推定の性能をシミュレーションによって評価する。

2 方法

2.1 シミュレーションの設定

シミュレーションの設定は西川(2019),p.107~p.114 の「3.5 シミュレーションによる推定方法の比較」[1]で行ったシミュレーションと同じ設定とし、無増悪生存時間を主要評価項目とした以下のような臨床試験を想定する。検査の回数はベースラインを含む 6

回で 12 週間隔とし、各検査時点のズレの許容幅を時点ごとに設定した。その時点に来院しなかった被験者の割合（欠測の確率）も考慮した（表 1）。

表 1 シミュレーションで設定した検査日程と欠測確率（西川(2019)p.109,図 3.7,一部改変[1]）

	治療開始	検査 1	検査 2	検査 3	検査 4	最終検査
計画された検査時点	0	12	24	36	48	60
各時点でのズレの許容幅	—	10~14	22~26	34~38	44~52	56~64
欠測の確率	0	0	0.1	0.1	0.1	0

無増悪生存時間の真の分布はワイブル分布（3 通り）を仮定し、尺度パラメータは形状パラメータごとに、60 週時点の無増悪生存率の期待値が 40%になるように数値計算により求めた。また、右側打ち切りが起こる時間の分布は一様分布を仮定し、最終検査時点より前に起こる右側打ち切りの割合（3 通り）ごとに、一様分布のサポートの長さを数値計算により求めた。パラメータの具体的な数値等は表 2、表 3、表 4 に示した。被験者数は 100、50、25、10 人の 4 通りで検討した。シミュレーションの試行回数は被験者数、無増悪生存時間の真の分布、右側打ち切りの割合の組合せ（合計 36 通り）ごとに 10000 回とした。

表 2 無増悪生存時間の真値と観測データの設定内容（西川(2019) p.107~p.110,一部改変[1]）

設定	内容
無増悪生存時間の真の分布	ワイブル分布を仮定し、形状パラメータは 0.667、1.00、1.50 の 3 通りとする。尺度パラメータは、60 週時点の無増悪生存率の期待値が 40%になるように各形状パラメータについて数値計算をして求めた。
右側打ち切りの割合	最終検査時点より前に起こる右側打ち切りの割合の期待値は 13%、25%、50%の 3 通りとする。また、無情報な打ち切りとする。
死亡の割合	17% 死亡の場合、イベント発現までの正確な時間が分かるとする。
試行回数	10000 回

このシミュレーションの設定は、西川(2019)[1]が実際に関わった臨床試験をもとにしている。この臨床試験は、Nishikawa ら(2003a)[14]によると治療の安全性を確認するために計画されたもので、研究デザインは単群であった。被験者数は 59 人でイベントを増悪、治療の中止または死亡のいずれか先に起きたものと定義した。ベースラインで

の検査の後、12 週間、24 週間、36 週間、48 週間、64 週間に検査を行い、それぞれ前後 2 週、前後 2 週、前後 2 週、前後 4 週、前後 4 週のずれを許容した。また、ベースラインを含め、それぞれの検査時点での欠測の割合は 0、0.012、0.093、0.087、0.049 および 0.132 であった。右側打ち切りの割合は 0.136 であり、死亡の割合は 0.172 であった。また、最終検査時点は、Nishikawa ら(2003a)[14]によると、「進行が遅い疾患の臨床試験では患者のほぼ 40%が、最後の検査でイベント未発現のまま観察終了となることがよくある」(Sarosdy et.al., (1998)[18]) ということ参考としていた。

2.2 シミュレーションの手順

西川(2019),p.107~p.114 の「3.5 シミュレーションによる推定方法の比較」[1]で行ったシミュレーションと同じ設定で以下のように観測データを生成し、シミュレーションを行う。

- 1 被験者数、ワイブル分布の形状パラメータ、右側打ち切りの割合を決めて以下のようにデータを生成し、いくつかの結果を記録する。
 - 1.1 ワイブル分布から被験者数分の無増悪生存時間のデータを生成する。
 - 1.2 無増悪生存時間のデータを確率 17%で死亡に振り分け、死亡を 1、増悪を 0 とする死亡振り分けデータを生成する。
 - 1.3 一様分布から被験者数分の右側打ち切りとなる時間のデータを生成する。
 - 1.4 無増悪生存時間のデータと右側打ち切りとなる時間のデータから被験者をイベント発現が観測されるか右側打ち切りになるかに振り分け、右側打ち切りとなる被験者を 1、右側打ち切りとならない被験者を 0 とする右側打ち切り振り分けデータを生成する。
 - 1.5 各被験者に対してズレを考慮した各検査の時間または欠測を表す検査日程のデータを生成する。
 - 1.6 検査日程のデータと死亡振り分けデータ、右側打ち切り振り分けデータから各被験者を正確な無増悪生存時間がわかる被験者、区間打ち切りとなる被験者、右側打ち切りとなる被験者に振り分ける。無増悪生存時間が最終検査時点より長い場合は、右側打ち切りとなる被験者に振り分ける。
 - 1.7 検査日程のデータと右側打ち切りとなる時間のデータから右側打ち切りとなる被験者の右側打ち切りが起こる検査時点を調べ、その時点を右側打ち切りが起こる時間とする。
 - 1.8 検査日程のデータと無増悪生存時間のデータから区間打ち切りとなる被験者の打ち切り区間を調べる。
 - 1.9 区間打ち切りとなる被験者に対し、それぞれ左端、中点、右端代入法を用いて 3 種類の無増悪生存時間のデータを作成する。
 - 1.10 代入法ごとに、区間打ち切りとなる被験者の無増悪生存時間、死亡に振り分けられた被験者の正確な無増悪生存時間（代入法によらず同一）、右側打ち切りとなる被験者の打ち切りまでの時間（代入法によらず同一）を用いて、イベントまたは打ち切りまでの時間と打ち切りかどうか分かるデータセットを 3 種類作成

する。

- 1.11 作成したデータセットごとに各時点のイベント数と打ち切り数を集計する。
- 1.12 集計したデータについて KM 法を行い、無増悪生存率を推定して記録する。
- 1.13 イベント発現時点ごとに GWSE を計算して記録する。
- 1.14 無増悪生存率と GWSE から両側 95%信頼区間および片側 97.5%信頼区間を構成し、仮定している真の無増悪生存率が含まれているかを調べて記録する。
- 1.15 二重対数変換を用いた場合の両側 95%信頼区間および片側 97.5%信頼区間についても、仮定している真の無増悪生存率が含まれているかを調べて記録する。
- 2 1 のデータの生成と結果の記録を 10000 回繰り返す。
- 3 評価指標を計算する。
 - 3.1 各時点の無増悪生存率の平均を計算する。
 - 3.2 各時点の GWSE の平均を計算する。
 - 3.3 各時点の無増悪生存率の SD を計算する。
 - 3.4 各時点の無増悪生存率の平均と仮定している真の無増悪生存率からバイアスを計算する。
 - 3.5 各時点の GWSE の平均から信頼区間幅を計算する。
 - 3.6 各時点の信頼区間に真の無増悪生存率が含まれている割合である被覆確率を計算する。
- 4 1～3 を被験者数、ワイブル分布の形状パラメータ、右側打ち切りの割合を変えて行う。

2.3 信頼区間の構成方法

$\varphi_{\alpha/2}$ を標準正規分布の上側 $\alpha/2$ 点とすると、各時点 t の両側 $100(1-\alpha)\%$ 信頼区間は以下のように求められる。

$$\hat{S}(t) \pm \varphi_{\alpha/2} \sqrt{\text{Var}(\hat{S}(t))}$$

ここに、 $\text{Var}(\hat{S}(t))$ は式 (1) により求める。 $\hat{S}(t)$ は 0～1 の値をとるが、 $\hat{S}(t)$ が 0 や 1 に近い場合や GWSE が大きい場合には GWSE (変換なし) の信頼区間は端点が 1 を超えたり 0 を下回ったりすることがある。二重対数変換を用いた信頼区間は以下のように求められる。

$$\exp \left[\log(\hat{S}(t)) \exp \left[\mp \varphi_{\alpha/2} \sqrt{\frac{\text{Var}(\hat{S}(t))}{\{\hat{S}(t) \log(\hat{S}(t))\}^2}} \right] \right]$$

二重対数変換後の信頼区間の端点は 0～1 に必ず入る。

また、上側信頼限界を U 、下側信頼限界を L とすると、 $S(t)$ は必ず 0～1 に入るので、上側信頼区間は $[L, 1]$ 、下側信頼区間は $[0, U]$ となる。各時点 t の上側 $100(1-\alpha)\%$ 信頼区間及び下側 $100(1-\alpha)\%$ 信頼区間は以下になる。

- ・ 上側 $100(1-\alpha)\%$ 信頼区間

$$\left[\hat{s}(t) - \varphi_{\alpha/2} \sqrt{\text{Var}(\hat{S}(t))}, 1 \right]$$

$$\left[\exp \left[\log(\hat{S}(t)) \exp \left[\varphi_{\alpha/2} \sqrt{\frac{\text{Var}(\hat{S}(t))}{\{\hat{S}(t) \log(\hat{S}(t))\}^2}} \right] \right], 1 \right]$$

- ・ 下側 $100(1-\alpha)\%$ 信頼区間

$$\left[0, \hat{s}(t) + \varphi_{\alpha/2} \sqrt{\text{Var}(\hat{S}(t))} \right]$$

$$\left[0, \exp \left[\log(\hat{S}(t)) \exp \left[-\varphi_{\alpha/2} \sqrt{\frac{\text{Var}(\hat{S}(t))}{\{\hat{S}(t) \log(\hat{S}(t))\}^2}} \right] \right] \right]$$

2.4 評価指標

評価指標として、GWSE と SD の差、両側 95% 信頼区間の被覆確率、片側 97.5% 信頼区間の被覆確率、無増悪生存率のバイアス及び信頼区間幅を用いる。

3 結果

3.1 パラメータの計算結果

ワイブル分布の形状パラメータごとの尺度パラメータは以下のようになった。

表 3 ワイブル分布の形状パラメータごとの尺度パラメータ

形状パラメータ	尺度パラメータ
0.6667	68.41
1.000	65.48
1.500	63.60

また、右側打ち切りの割合ごとの一様分布のサポートの長さは以下のようになった。

表 4 右側打ち切りの割合ごとの一様分布のサポートの長さ

無増悪生存時間 の真の分布	右側打ち切りの割合の期待値		
	13%	25%	50%
ワイブル (0.6667,68,41)	275	143	72
ワイブル (1.000,65.48)	302	158	79
ワイブル (1.500,63.60)	332	173	87

シミュレーションで設定するワイブル分布の尺度パラメータと一様分布のサポートの長さについて、数値計算の結果と西川(2019),p.110[1]の値を比較し、シミュレーションの設定について確認した。尺度パラメータ（表 3）は西川(2019),p.110[1]での尺度パラメータの値と等しくなった。また、一様分布のサポートの長さ（表 4）については、西川(2019),p.110[1]でのサポートの長さとの差は 0～2 であり、ほとんど同じ値となった。

3.2 シミュレーションの結果

被覆確率の計算には GWSE（変換なし）の信頼区間と二重対数変換後の信頼区間を用いた。右端代入法、中点代入法を用いた場合は、二重対数変換後の信頼区間を用いた被覆確率の方が名目の信頼水準に近い値であった。左端代入法を用いた場合、試験の前半は変換なしの信頼区間を用いた被覆確率の方が名目の信頼水準に近く、後半からは二重対数変換後の信頼区間を用いた被覆確率の方が名目の信頼水準に近い値であった。変換なしの信頼区間を用いた被覆確率の方が二重対数変換後の信頼区間を用いた被覆確率より高い区間は被験者数が多いほど長かった。また、2つの被覆確率の差は形状パラメータが高いほど大きかった。GWSE（変換なし）の信頼区間の被覆確率と二重対数変換後の信頼区間の被覆確率は各時点の被覆確率の差はあるが、似た形状をしておりパラメータの組み合わせによる特徴も似ていた。これ以降の結果では二重対数変換後の信頼区間を用いた被覆確率についてまとめる。

図 4 から図 7 はそれぞれのパラメータごとに両側 95%信頼区間の被覆確率、下側 97.5%信頼区間の被覆確率、上側 97.5%信頼区間の被覆確率及び GWSE と SD の差を比較したグラフである。Nishikawa ら(2003a)[14]では多くの場合、中点代入法が平均二乗誤差を小さくしていた。そのため、代入法の比較以外の図では中点代入法を用いた場合の図を示した。左端代入法、右端代入法を用いた場合の図は付録に掲載する。表 5 から表 8 はパラメータの組み合わせごとにベースラインから 64 週の各時点における両側 95%信頼区間の被覆確率が名目の信頼水準の前後 5%及び 1%以内である時点の割合、上側下側両方の片側 97.5%信頼区間の被覆確率が名目の信頼水準の前後 2.5%及び 0.5%以内である時点の割合をまとめたものであり、表 7 は最終検査時点の被覆確率を評価しまとめたものである。また、表 10、表 11、表 12 はそれぞれ代入法ごとに被験者

生存時間解析における Greenwood 式標準誤差推定の性能評価 (榎本, 西川(哲), 西川(正))
数ごとのバイアス、信頼区間幅及び被覆確率を比較した数値表である。以下では、シミュレーションの結果から分かることをそれぞれの評価指標ごとにまとめる。

・ GWSE について

被験者数が少ないほど GWSE は SD を過小評価していた (図 4)。しかし、被験者数が 100 人の場合は右端代入法を用いた場合を除いて SD を正しく推定できていた (図 4)。右端代入法を用いた場合は、最終検査時点付近から GWSE は SD を過小評価していた (図 5)。また、右側打ち切りの割合が高い場合も GWSE は SD を過小評価していた (図 7) が、被験者数が多いほどその影響は少なくなっていた。

・ 被覆確率について

すべての時点で被覆確率が名目の信頼水準を満たしているパラメータの組み合わせはなかった (表 7、表 8)。しかし、大部分の被覆確率が名目の信頼水準に近い (両側 95%信頼区間の被覆確率では名目の信頼水準の前後 5%以内、片側 97.5%信頼区間の被覆確率では名目の信頼水準の前後 2.5%以内) パラメータの組み合わせはあった (表 5、表 6)。全体的に被覆確率は被験者数、形状パラメータ、右側打ち切りの割合が低いほど名目の信頼水準に近い傾向がある (図 4、図 6、図 7)。

左端代入法、中点代入法を用いた場合、下側 97.5%信頼区間の被覆確率が低く、上側 97.5%信頼区間の被覆確率が高い傾向があり、左端代入法で顕著であった (図 5)。また、右端代入法を用いた場合、試験の前半は上側 97.5%信頼区間の被覆確率が低く、後半は下側 97.5%信頼区間の被覆確率が低い傾向があった。しかし、最終検査時点だけをみると多くの場合被覆確率は名目の信頼水準に近く、1 点代入法の中では左端代入法が最良であった (表 7)。

代入法ごとに大部分の片側 97.5%信頼区間の被覆確率が名目の信頼水準に近くなるパラメータの組み合わせがあった。左端代入法を用いた場合では、上側 97.5%信頼区間の被覆確率は被験者数が 25 人以上の場合、1 回目の検査時点から最終検査時点まで名目の信頼水準より高くなっている。

中点代入法を用いた場合では、上側 97.5%信頼区間の被覆確率は被験者数が 50 人以上の場合、または被験者数が 25 人で形状パラメータが低い場合、6 週から最終検査時点まで名目の信頼水準より高くなっている。下側 97.5%信頼区間の被覆確率は、被験者数が 25 人以上で右側打ち切りの割合が 0.13 の場合、各検査時点から数週後までは名目の信頼水準に近い値となっている。

右端代入法を用いた場合では上側 97.5%信頼区間の被覆確率は被験者数が 25 人の場合、形状パラメータが低いとき 1 回目の検査時点以降は名目の信頼水準に近い値となっている。また、形状パラメータが 1.5 のとき 2 回目の検査時点以降は名目の信頼水準に近い値となっている。被験者数が 50 人の場合、右側打ち切りの割合が 0.5 ならば 1 回目の検査時点以降は名目の信頼水準に近い値となっている。下側 97.5%信頼区間の被覆確率は右側打ち切りの割合が低い場合、ベースラインから最終検査時点まで名目の信頼水準に近い値であり、被験者数が多い場合は名目の信頼水準より高い時点が多くなっ

ている。

それぞれのパラメータごとに被覆確率を比較した。被験者数が多いほど左端代入法を用いた場合の下側 97.5%信頼区間の被覆確率と右端代入法を用いた場合の上側 97.5%信頼区間の被覆確率にみられる各検査日の間での被覆確率の増減の幅が大きかった。また、形状パラメータが大きいほど左端代入法を用いた場合の下側 97.5%信頼区間の被覆確率、中点代入法を用いた場合の上側 97.5%信頼区間の被覆確率、右端代入法を用いた場合の上側 97.5%信頼区間の被覆確率は低くなった。中点代入法を用いた場合の上側 97.5%信頼区間の被覆確率、右端代入法を用いた場合の上側 97.5%信頼区間の被覆確率は被験者数が少ないほど形状パラメータの影響が大きかった。右端代入法を用いた場合の下側 97.5%信頼区間の被覆確率は形状パラメータが大きいほど高くなった。右側打ち切りの割合が高いほど中点代入法、右端代入法を用いた場合の下側 97.5%信頼区間の被覆確率は低くなっている。左端代入法を用いた場合の下側 97.5%信頼区間の被覆確率は被験者数が少ない場合は右側打ち切りの割合が高いほど低くなるが、被験者数が多い場合はわずかに高くなる。

表 5 両側 95%信頼区間の被覆確率が名目の信頼水準の前後 5%以内である時点の割合

		f=2/3			f=1.0			f=1.5		
		d=0.13	d=0.25	d=0.5	d=0.13	d=0.25	d=0.5	d=0.13	d=0.25	d=0.5
L	n=10	71%	74%	42%	58%	58%	28%	34%	37%	3%
	n=25	71%	71%	75%	51%	51%	58%	18%	22%	42%
	n=50	54%	54%	68%	22%	31%	45%	15%	15%	20%
	n=100	40%	42%	48%	15%	17%	23%	12%	14%	14%
M	n=10	85%	82%	46%	69%	68%	20%	54%	49%	15%
	n=25	89%	86%	60%	85%	85%	65%	74%	71%	58%
	n=50	86%	83%	40%	83%	80%	49%	80%	75%	55%
	n=100	80%	68%	15%	77%	63%	26%	65%	60%	32%
R	n=10	78%	75%	45%	63%	60%	31%	46%	43%	20%
	n=25	82%	80%	60%	82%	80%	74%	68%	68%	62%
	n=50	75%	77%	49%	68%	74%	72%	62%	75%	77%
	n=100	60%	68%	34%	43%	55%	48%	26%	46%	66%

L : 左端代入法、M : 中点代入法、R : 右端代入法、n : 被験者数、f : 形状パラメータ、d : 右側打ち切りの割合

表 6 両方の片側 97.5%信頼区間の被覆確率が名目の信頼水準の前後 2.5%以内である時点の割合

		f=2/3			f=1.0			f=1.5		
		d=0.13	d=0.25	d=0.5	d=0.13	d=0.25	d=0.5	d=0.13	d=0.25	d=0.5
L	n=10	25%	14%	2%	11%	8%	0%	9%	6%	0%
	n=25	17%	17%	2%	12%	12%	12%	11%	11%	11%
	n=50	14%	15%	29%	11%	12%	15%	9%	9%	11%
	n=100	12%	14%	17%	9%	9%	12%	6%	6%	9%
M	n=10	57%	15%	2%	34%	23%	3%	23%	18%	2%
	n=25	63%	42%	5%	57%	40%	9%	42%	32%	6%
	n=50	52%	26%	6%	51%	34%	9%	46%	35%	14%
	n=100	42%	17%	5%	42%	25%	9%	40%	32%	12%
R	n=10	71%	66%	17%	57%	52%	6%	40%	37%	14%
	n=25	78%	75%	26%	74%	75%	43%	57%	62%	54%
	n=50	55%	63%	26%	42%	57%	40%	25%	42%	60%
	n=100	32%	49%	17%	12%	35%	28%	14%	18%	43%

L : 左端代入法、M : 中点代入法、R : 右端代入法、n : 被験者数、f : 形状パラメータ、d : 右側打ち切りの割合

表 7 両側 95%信頼区間の被覆確率が名目の信頼水準の前後 1%以内である時点の割合

		f=2/3			f=1.0			f=1.5		
		d=0.13	d=0.25	d=0.5	d=0.13	d=0.25	d=0.5	d=0.13	d=0.25	d=0.5
L	n=10	45%	26%	2%	20%	9%	0%	12%	9%	0%
	n=25	20%	18%	2%	14%	14%	0%	11%	11%	0%
	n=50	14%	15%	0%	11%	12%	9%	8%	9%	11%
	n=100	12%	12%	9%	8%	9%	11%	6%	6%	6%
M	n=10	62%	34%	0%	45%	25%	0%	37%	14%	0%
	n=25	62%	38%	6%	52%	40%	11%	38%	28%	9%
	n=50	52%	28%	6%	46%	26%	3%	40%	28%	11%
	n=100	40%	15%	5%	40%	23%	8%	37%	26%	6%
R	n=10	26%	48%	3%	17%	32%	0%	11%	15%	0%
	n=25	66%	62%	20%	52%	48%	25%	38%	38%	37%
	n=50	55%	60%	28%	37%	52%	31%	15%	29%	46%
	n=100	34%	49%	17%	11%	35%	28%	11%	12%	34%

L : 左端代入法、M : 中点代入法、R : 右端代入法、n : 被験者数、f : 形状パラメータ、d : 右側打ち切りの割合

表 8 両方の片側 97.5%信頼区間の被覆確率が名目の信頼水準の前後 0.5%以内である時点の割合

		f=2/3			f=1.0			f=1.5		
		d=0.13	d=0.25	d=0.5	d=0.13	d=0.25	d=0.5	d=0.13	d=0.25	d=0.5
L	n=10	0%	0%	0%	0%	0%	0%	0%	0%	0%
	n=25	5%	6%	0%	3%	5%	0%	3%	3%	0%
	n=50	5%	5%	0%	3%	3%	2%	2%	2%	3%
	n=100	3%	5%	0%	3%	3%	3%	2%	3%	3%
M	n=10	0%	0%	0%	2%	0%	0%	0%	0%	0%
	n=25	0%	0%	0%	2%	0%	0%	3%	2%	0%
	n=50	2%	2%	0%	8%	0%	0%	3%	2%	0%
	n=100	9%	0%	2%	9%	5%	0%	8%	5%	0%
R	n=10	0%	2%	0%	0%	2%	0%	6%	6%	0%
	n=25	14%	3%	0%	2%	5%	0%	2%	5%	0%
	n=50	20%	15%	2%	3%	8%	3%	0%	2%	8%
	n=100	6%	8%	0%	2%	8%	3%	0%	3%	5%

L : 左端代入法、M : 中点代入法、R : 右端代入法、n : 被験者数、f : 形状パラメータ、d : 右側打ち切りの割合

表 9 最終検査時点 (60 週) の片側 97.5%信頼区間の被覆確率の評価

		f=2/3			f=1.0			f=1.5		
		d=0.13	d=0.25	d=0.5	d=0.13	d=0.25	d=0.5	d=0.13	d=0.25	d=0.5
L	n=10	○	○	△	○	○	△	○	△	▲
	n=25	○	○▼	△	○	○	○	○	○	○▲
	n=50	○	○▼	○▼	○	○▼	●	○	○▼	○▼
	n=100	●	●	○▼	●	●	●	●	●	●
M	n=10	△	△	△	△	△	△	△	△	△
	n=25	○	○	△	○	○	△	○	○	△
	n=50	○	△	△	○	△	△	○	○	△
	n=100	○	△	△	○	△	△	○	△	△
R	n=10	△	△	△	○	△	△	○	△	△
	n=25	○▼	○	△	●	○	△	○	●	△
	n=50	○	○	△	○▼	○	△	○	○▼	△
	n=100	●	△	△	○	○	△	▽	○	△

● : 両方とも名目の信頼水準の前後 0.5%以内、▼ : 下側が名目の信頼水準の前後 0.5%以内、▲ : 上側が名目の信頼水準の前後 0.5%以内、○ : 両方とも名目の信頼水準の前後 2.5%以内、▽ : 下側が名目の信頼水準の前後 2.5%以内、△ : 上側が名目の信頼水準の前後 2.5%以内、L : 左端代入法、M : 中点代入法、R : 右端代入法、n : 被験者数、f : 形状パラメータ、d : 右側打ち切りの割合

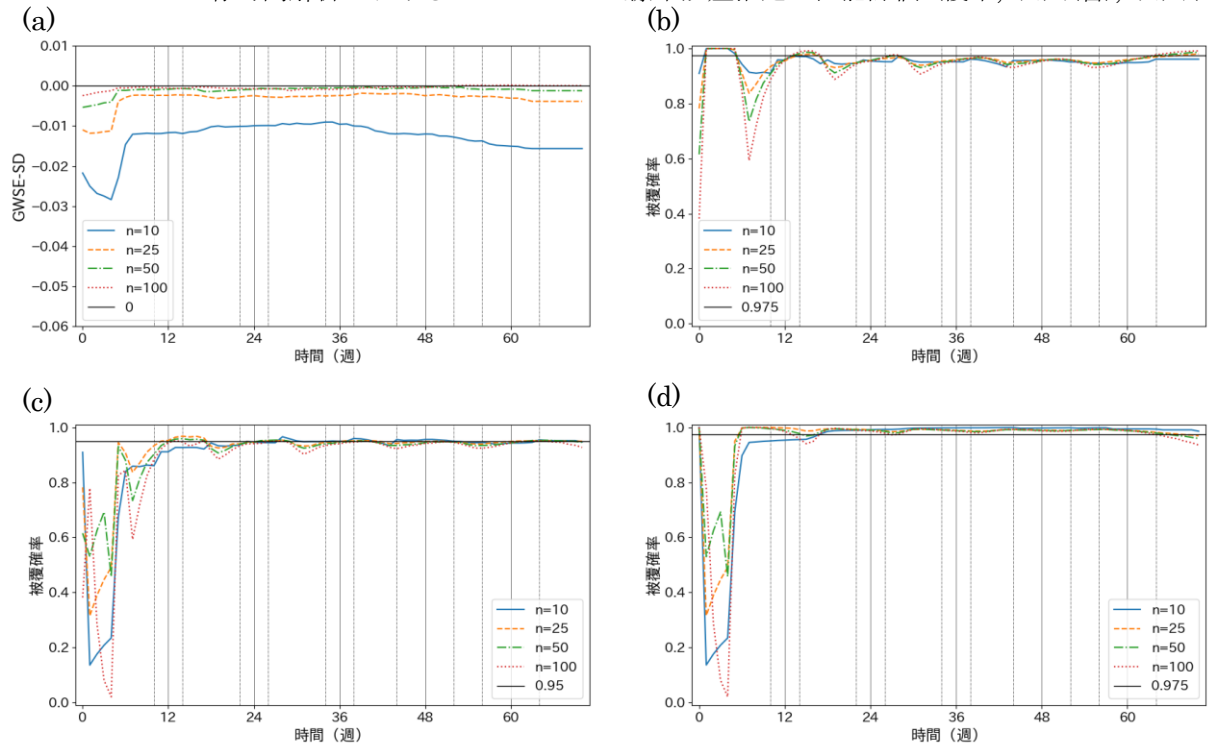


図 4 被験者数ごとの GWSE と SD の差及び被覆確率の比較 (中点代入法、形状パラメータ=2/3、右側打ち切りの割合=0.13、 n : 被験者数、(a) GWSE と SD の差、(b) 下側 97.5%信頼区間の被覆確率、(c) 両側 95%信頼区間の被覆確率、(d) 上側 97.5%信頼区間の被覆確率)

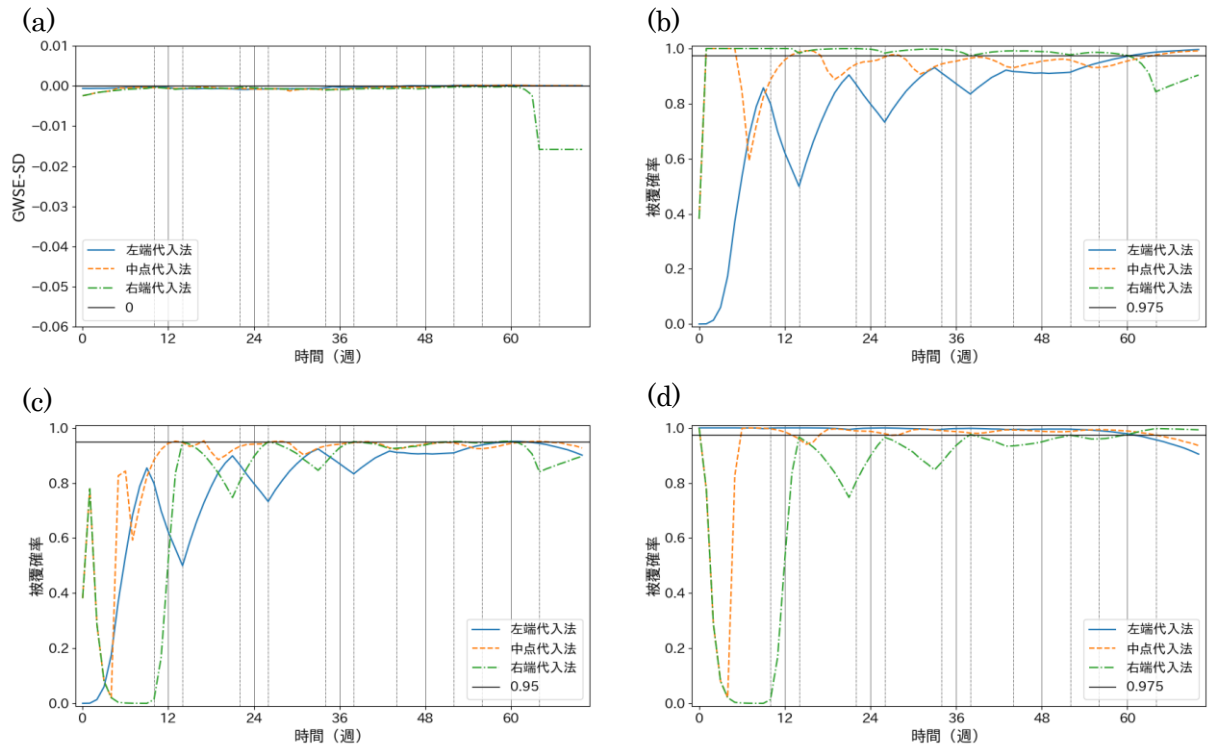


図 5 代入法ごとの GWSE と SD の差及び被覆確率の比較 (被験者数=100、形状パラメータ=2/3、右側打ち切りの割合=0.13、(a) GWSE と SD の差、(b) 下側 97.5%信頼区間の被覆確率、(c) 両側 95%信頼区間の被覆確率、(d) 上側 97.5%信頼区間の被覆確率)

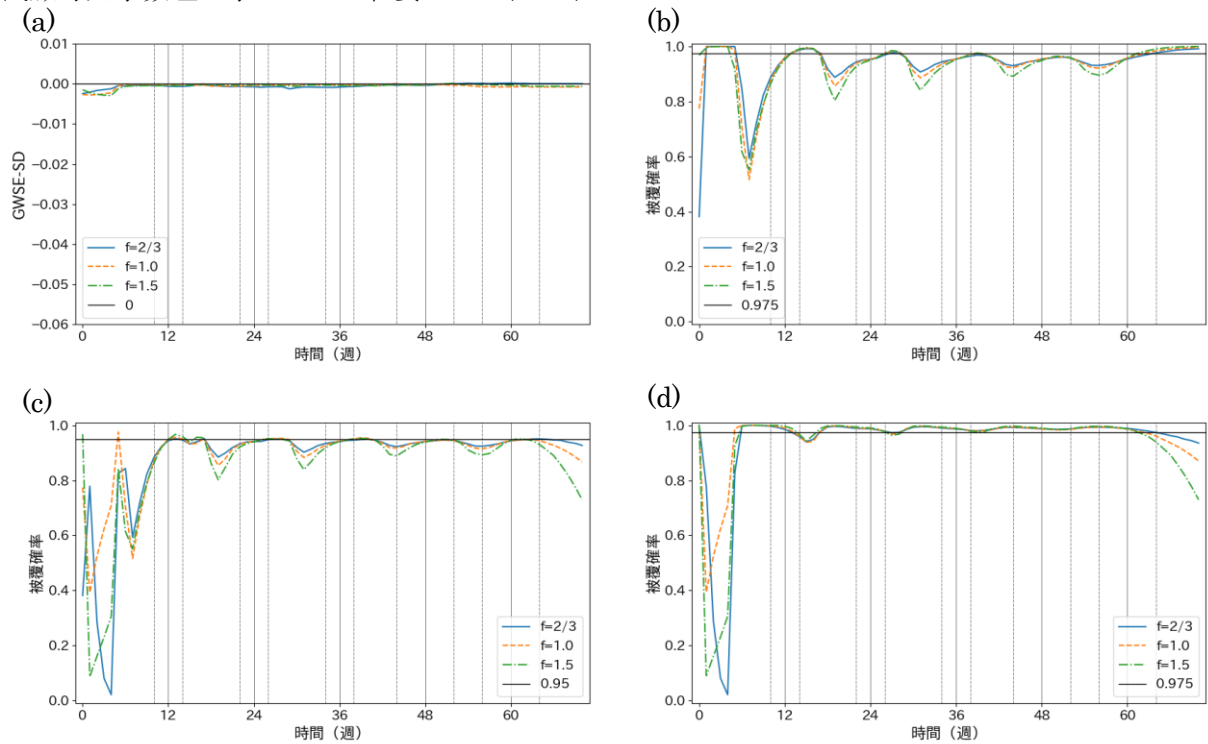


図 6 形状パラメータごとの GWSE と SD の差及び被覆確率の比較 (中点代入法、被験者数=100、右側打ち切りの割合=0.13、 f : 形状パラメータ、(a) GWSE と SD の差、(b) 下側 97.5%信頼区間の被覆確率、(c) 両側 95%信頼区間の被覆確率、(d) 上側 97.5%信頼区間の被覆確率)

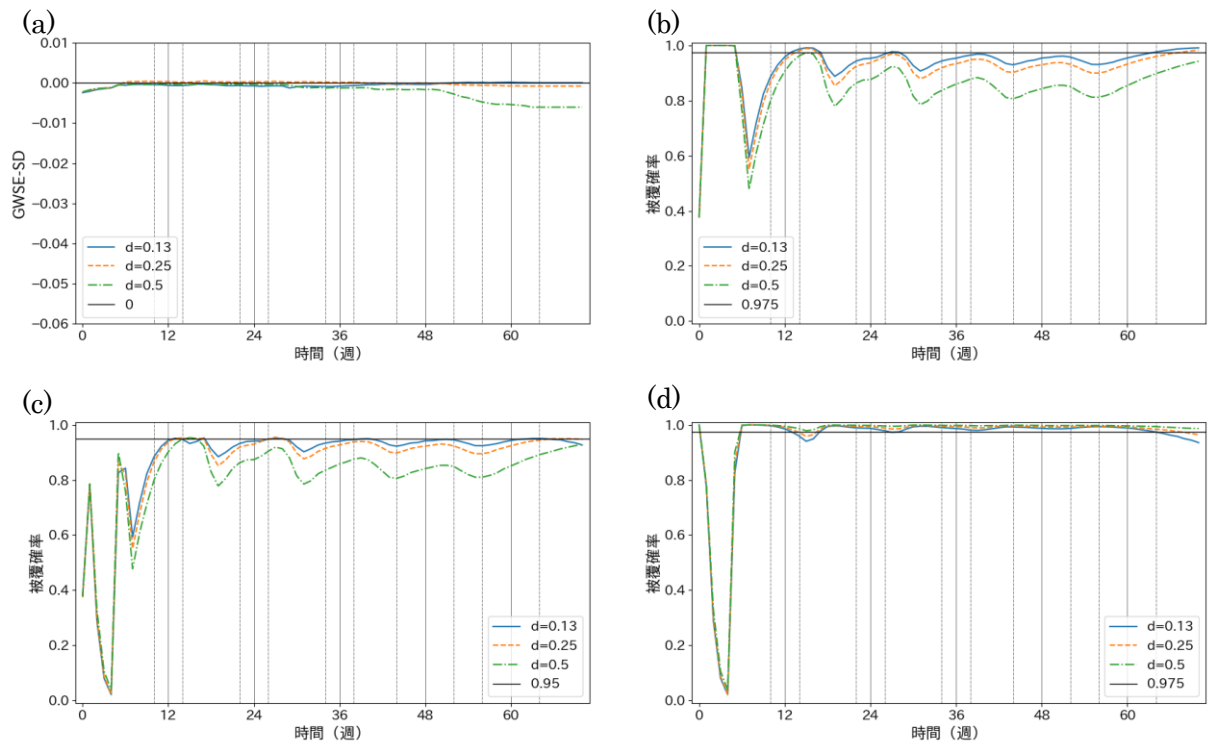


図 7 右側打ち切りの割合ごとの GWSE と SD の差及び被覆確率の比較 (中点代入法、被験者数=100、形状パラメータ=2/3、 d : 右側打ち切りの割合、(a) GWSE と SD の差、(b) 下側 97.5%信頼区間の被覆確率、(c) 両側 95%信頼区間の被覆確率、(d) 上側 97.5%信頼区間の被覆確率)

・無増悪生存率のバイアスについて

被験者数が 10 人の場合、右側打ち切りの割合が 0.5 のときバイアスは正の方向にわずかに大きくなる。また、右端代入法を用いた場合は被験者数が多いほど最終検査時点後のバイアスが負の方向に大きくなる。それ以外の場合は、被験者数によるバイアスの違いは無かった (表 10、表 11、表 12)。多くの場合、バイアスは右端代入法、中点代入法、左端代入法の順に負の方向に大きくなる (表 10、表 11、表 12)。形状パラメータによるバイアスの違いは、左端代入法を用いた場合では 1 回目の検査時点付近から最終検査時点付近まで形状パラメータが大きいほど負の方向にバイアスが大きくなっている。また、右端代入法を用いた場合では 1 回目の検査時点付近から最終検査時点付近まで形状パラメータが大きいほど正の方向にバイアスが大きくなっている。右側打ち切りの割合によるバイアスの違いは、中点代入法、右端代入法を用いた場合、右側打ち切りの割合が高いほど負の方向にバイアスが大きくなっている。

・信頼区間幅について

右端代入法、中点代入法を用いた場合、ベースラインから 1 回目の検査時点までに代入される時点から大幅に信頼区間幅は広くなる。また、代入法によらず被験者数が少ないほど信頼区間幅は広くなる。例えば、形状パラメータが 2/3、右側打ち切りの割合が 0.13 の場合、被験者数によって最終検査時点での信頼区間幅は 0.2 から 0.5 程度まで変化する (表 10、表 11、表 12)。右側打ち切りの割合が大きいほど、信頼区間幅は広くなる。一方、形状パラメータの値が大きいほど、早い時点では信頼区間幅は狭くなるが、遅い時点では形状パラメータの値による差はない。

表 10 左端代入法を用いた場合の被験者数ごとのバイアス、信頼区間幅及び被覆確率の比較

		t	6	12	18	24	30	36	42	48	54	60
L	n=10	無増悪生存率	0.74723	0.66047	0.60988	0.55436	0.52066	0.47874	0.45286	0.42073	0.40182	0.39846
		バイアス	-0.0736	-0.0705	-0.0534	-0.0537	-0.0408	-0.0424	-0.0327	-0.0333	-0.0238	-0.0015
		信頼区間幅	0.5044	0.5405	0.5522	0.5580	0.5590	0.5563	0.5524	0.5460	0.5410	0.5399
		両側95%信頼区間の被覆確率	0.8529	0.8996	0.9246	0.9156	0.9447	0.9407	0.9259	0.9484	0.9458	0.9508
		下側97.5%信頼区間の被覆確率	0.9099	0.9173	0.9335	0.9196	0.9469	0.9415	0.9262	0.9507	0.9480	0.9606
		上側97.5%信頼区間の被覆確率	0.9430	0.9823	0.9911	0.9960	0.9978	0.9992	0.9997	0.9977	0.9978	0.9902
	n=25	無増悪生存率	0.7481	0.6616	0.6118	0.5572	0.5244	0.4826	0.4569	0.4245	0.4057	0.4017
		バイアス	-0.0728	-0.0693	-0.0514	-0.0509	-0.0371	-0.0385	-0.0287	-0.0295	-0.0200	0.0017
		信頼区間幅	0.3373	0.3640	0.3741	0.3812	0.3838	0.3849	0.3847	0.3831	0.3816	0.3819
		両側95%信頼区間の被覆確率	0.8444	0.8282	0.8890	0.9008	0.9258	0.9251	0.9331	0.9325	0.9415	0.9515
		下側97.5%信頼区間の被覆確率	0.8452	0.8284	0.8919	0.9048	0.9309	0.9306	0.9394	0.9384	0.9494	0.9675
		上側97.5%信頼区間の被覆確率	0.9992	0.9998	0.9971	0.9960	0.9949	0.9945	0.9937	0.9941	0.9921	0.9840
	n=50	無増悪生存率	0.7469	0.6606	0.6110	0.5557	0.5226	0.4805	0.4550	0.4231	0.4041	0.4003
		バイアス	-0.0739	-0.0704	-0.0523	-0.0524	-0.0389	-0.0406	-0.0306	-0.0309	-0.0215	0.0003
		信頼区間幅	0.2408	0.2619	0.2702	0.2764	0.2789	0.2803	0.2805	0.2799	0.2792	0.2796
		両側95%信頼区間の被覆確率	0.7299	0.7850	0.8532	0.8643	0.9010	0.9039	0.9239	0.9219	0.9339	0.9505
		下側97.5%信頼区間の被覆確率	0.7305	0.7856	0.8558	0.8669	0.9057	0.9080	0.9300	0.9272	0.9427	0.9694
		上側97.5%信頼区間の被覆確率	0.9994	0.9994	0.9974	0.9974	0.9953	0.9959	0.9939	0.9947	0.9912	0.9811
	n=100	無増悪生存率	0.7477	0.6609	0.6112	0.5561	0.5231	0.4812	0.4558	0.4239	0.4047	0.4010
		バイアス	-0.0732	-0.0701	-0.0520	-0.0520	-0.0384	-0.0398	-0.0298	-0.0302	-0.0210	0.0010
		信頼区間幅	0.1705	0.1866	0.1930	0.1978	0.1997	0.2011	0.2014	0.2012	0.2008	0.2011
		両側95%信頼区間の被覆確率	0.5343	0.6224	0.7864	0.7979	0.8694	0.8701	0.9036	0.9065	0.9260	0.9506
		下側97.5%信頼区間の被覆確率	0.5345	0.6225	0.7873	0.7989	0.8723	0.8726	0.9080	0.9114	0.9328	0.9719
		上側97.5%信頼区間の被覆確率	0.9998	0.9999	0.9991	0.9990	0.9971	0.9975	0.9956	0.9951	0.9932	0.9787

形状パラメータ=2/3、右側打ち切りの割合=0.13、L：左端代入法、n：被験者数

表 11 中点代入法を用いた場合の被験者数ごとのバイアス、信頼区間幅及び被覆確率の比較

		t	6	12	18	24	30	36	42	48	54	60
M	n=10	無増悪生存率	0.7880	0.7230	0.6389	0.5963	0.5366	0.5086	0.4689	0.4402	0.4042	0.3875
		バイアス	-0.0329	-0.0079	-0.0243	-0.0118	-0.0249	-0.0125	-0.0167	-0.0138	-0.0215	-0.0125
		信頼区間幅	0.4850	0.5224	0.5512	0.5580	0.5626	0.5615	0.5572	0.5522	0.5432	0.5379
		両側95%信頼区間の被覆確率	0.8398	0.9111	0.9434	0.9458	0.9508	0.9498	0.9420	0.9558	0.9446	0.9432
		下側97.5%信頼区間の被覆確率	0.9428	0.9577	0.9584	0.9545	0.9539	0.9518	0.9430	0.9584	0.9473	0.9490
		上側97.5%信頼区間の被覆確率	0.8970	0.9534	0.9850	0.9913	0.9969	0.9980	0.9990	0.9974	0.9973	0.9942
	n=25	無増悪生存率	0.7879	0.7252	0.6417	0.5986	0.5397	0.5116	0.4725	0.4435	0.4082	0.3905
		バイアス	-0.0329	-0.0057	-0.0216	-0.0094	-0.0217	-0.0095	-0.0131	-0.0105	-0.0174	-0.0095
		信頼区間幅	0.3267	0.3504	0.3726	0.3795	0.3856	0.3866	0.3870	0.3856	0.3829	0.3816
		両側95%信頼区間の被覆確率	0.9074	0.9525	0.9313	0.9460	0.9357	0.9482	0.9475	0.9483	0.9416	0.9482
		下側97.5%信頼区間の被覆確率	0.9100	0.9550	0.9374	0.9560	0.9433	0.9599	0.9573	0.9588	0.9504	0.9594
		上側97.5%信頼区間の被覆確率	0.9974	0.9975	0.9939	0.9900	0.9924	0.9883	0.9902	0.9895	0.9912	0.9888
	n=50	無増悪生存率	0.7876	0.7235	0.6401	0.5975	0.5387	0.5103	0.4708	0.4413	0.4066	0.3888
		バイアス	-0.0332	-0.0075	-0.0231	-0.0106	-0.0228	-0.0107	-0.0148	-0.0128	-0.0190	-0.0112
		信頼区間幅	0.2319	0.2510	0.2688	0.2745	0.2800	0.2811	0.2820	0.2815	0.2800	0.2796
		両側95%信頼区間の被覆確率	0.8808	0.9488	0.9231	0.9463	0.9326	0.9451	0.9406	0.9445	0.9380	0.9431
		下側97.5%信頼区間の被覆確率	0.8832	0.9601	0.9306	0.9566	0.9398	0.9572	0.9516	0.9559	0.9477	0.9554
		上側97.5%信頼区間の被覆確率	0.9976	0.9887	0.9925	0.9897	0.9928	0.9879	0.9890	0.9886	0.9903	0.9877
	n=100	無増悪生存率	0.7884	0.7243	0.6403	0.5978	0.5389	0.5105	0.4711	0.4423	0.4072	0.3896
		バイアス	-0.0324	-0.0067	-0.0229	-0.0103	-0.0226	-0.0106	-0.0145	-0.0117	-0.0185	-0.0104
		信頼区間幅	0.1636	0.1781	0.1918	0.1962	0.2005	0.2015	0.2024	0.2022	0.2014	0.2012
		両側95%信頼区間の被覆確率	0.8430	0.9442	0.9127	0.9409	0.9203	0.9404	0.9383	0.9420	0.9326	0.9434
		下側97.5%信頼区間の被覆確率	0.8448	0.9588	0.9188	0.9535	0.9274	0.9539	0.9488	0.9544	0.9409	0.9546
		上側97.5%信頼区間の被覆確率	0.9982	0.9854	0.9939	0.9874	0.9929	0.9865	0.9895	0.9876	0.9917	0.9888

形状パラメータ=2/3、右側打ち切りの割合=0.13、M：中点代入法、n：被験者数

表 12 右端代入法を用いた場合の被験者数ごとのバイアス、信頼区間幅及び被覆確率の比較

		t	6	12	18	24	30	36	42	48	54	60
R	n=10	無増悪生存率	0.96653	0.81968	0.70743	0.64588	0.59552	0.54551	0.50569	0.47137	0.43628	0.39966
		バイアス	0.1457	0.0887	0.0442	0.0378	0.0341	0.0244	0.0201	0.0174	0.0106	-0.0003
		信頼区間幅	0.1671	0.4602	0.5324	0.5526	0.5610	0.5647	0.5638	0.5598	0.5531	0.5441
		両側95%信頼区間の被覆確率	0.2821	0.8446	0.9441	0.9621	0.9698	0.9640	0.9641	0.9668	0.9566	0.9389
		下側97.5%信頼区間の被覆確率	1.0000	0.9932	0.9844	0.9791	0.9786	0.9683	0.9661	0.9716	0.9618	0.9470
		上側97.5%信頼区間の被覆確率	0.2821	0.8514	0.9597	0.9830	0.9912	0.9957	0.9980	0.9952	0.9948	0.9919
	n=25	無増悪生存率	0.9663	0.8193	0.7099	0.6481	0.5985	0.5486	0.5083	0.4744	0.4396	0.4028
		バイアス	0.1454	0.0883	0.0466	0.0401	0.0371	0.0275	0.0227	0.0204	0.0139	0.0028
		信頼区間幅	0.1436	0.3126	0.3579	0.3733	0.3819	0.3870	0.3889	0.3890	0.3874	0.3913
		両側95%信頼区間の被覆確率	0.5642	0.9610	0.9626	0.9559	0.9554	0.9566	0.9568	0.9574	0.9563	0.9533
		下側97.5%信頼区間の被覆確率	1.0000	0.9979	0.9884	0.9854	0.9844	0.9817	0.9800	0.9794	0.9758	0.9706
		上側97.5%信頼区間の被覆確率	0.5642	0.9631	0.9742	0.9705	0.9710	0.9749	0.9768	0.9780	0.9805	0.9827
	n=50	無増悪生存率	0.9664	0.8200	0.7081	0.6468	0.5969	0.5473	0.5073	0.4733	0.4375	0.4007
		バイアス	0.1456	0.0890	0.0449	0.0388	0.0354	0.0262	0.0217	0.0193	0.0118	0.0007
		信頼区間幅	0.1120	0.2205	0.2568	0.2692	0.2764	0.2810	0.2829	0.2835	0.2829	0.2879
		両側95%信頼区間の被覆確率	0.1508	0.7881	0.9301	0.9371	0.9396	0.9497	0.9489	0.9515	0.9522	0.9496
		下側97.5%信頼区間の被覆確率	1.0000	0.9996	0.9944	0.9921	0.9902	0.9869	0.9846	0.9829	0.9761	0.9695
		上側97.5%信頼区間の被覆確率	0.1508	0.7885	0.9357	0.9450	0.9494	0.9628	0.9643	0.9686	0.9761	0.9801
	n=100	無増悪生存率	0.9660	0.8202	0.7090	0.6471	0.5971	0.5474	0.5074	0.4736	0.4383	0.4012
		バイアス	0.1452	0.0892	0.0457	0.0390	0.0356	0.0263	0.0218	0.0196	0.0126	0.0012
		信頼区間幅	0.0794	0.1551	0.1824	0.1920	0.1975	0.2012	0.2028	0.2034	0.2032	0.2075
		両側95%信頼区間の被覆確率	0.0012	0.5206	0.8710	0.8957	0.9070	0.9275	0.9337	0.9378	0.9476	0.9511
		下側97.5%信頼区間の被覆確率	1.0000	1.0000	0.9983	0.9976	0.9955	0.9917	0.9893	0.9888	0.9837	0.9750
		上側97.5%信頼区間の被覆確率	0.0012	0.5206	0.8727	0.8981	0.9115	0.9358	0.9444	0.9490	0.9639	0.9761

形状パラメータ=2/3、右側打ち切りの割合=0.13、R：右端代入法、n：被験者数

4 考察

表 7、表 8 からどの信頼区間の被覆確率もすべての時点で名目の信頼水準を満たしているパラメータの組み合わせが無いことが分かる。中点代入法や右端代入法を用いた場合は、両側 95%信頼区間の被覆確率が名目の信頼水準の前後 1%以内である時点の割合が 6 割程度になるパラメータの組み合わせがあった。しかし、両方の片側 97.5%信頼区間の被覆確率が名目の信頼水準の前後 0.5%以内である時点の割合は高くても 20%であり、ほとんどの場合は数%であった。

表 5、表 6 から両側 95%信頼区間の被覆確率が名目の信頼水準の前後 5%以内である区間および片側 97.5%信頼区間の被覆確率が名目の信頼水準の前後 2.5%以内である区間が大部分を占めるパラメータの組み合わせは存在することが分かる。形状パラメータや右側打ち切りの割合は低い方が名目の信頼水準に近かった。また、被験者数は少ない方が名目の信頼水準に近いが、10 人よりも 25 人の方が良い結果であった (図 4、表 5、表 6)。被験者数が少ないほど名目の信頼水準に近いのは、被験者数によってバイアスが変わらないが、被験者数が少ないほど GWSE は大きくなり信頼区間幅が広くなることで、信頼区間幅が被覆確率に与える影響が大きくなることが原因だと考えられる (表 10、表 11、表 12)。また、被験者数が 10 人より 25 人の方が名目の信頼水準に近いのは、被験者数が 25 人の場合と比べ 10 人の場合は形状パラメータや右側打ち切りの割合による被覆確率への影響が大きいためと考えられる。

代入法ごとに被覆確率に異なる特徴が見られた。左端代入法を用いた場合、被験者数が少なく、形状パラメータ、右側打ち切りの割合が小さければ両側 95%信頼区間の被覆確率が名目の信頼水準の前後 5%以内である時点の割合は 7 割程度ある。しかし、負の方向にバイアスが大きい下側 97.5%信頼区間の被覆確率が低く、上側 97.5%信頼区間の被覆確率が高くなり、両方の片側 97.5%信頼区間の被覆確率が名目の信頼水準の前後 2.5%以内である時点の割合は低くなっている。中点代入法を用いた場合、右側打ち切りの割合が低ければ両側 95%信頼区間の被覆確率が名目の信頼水準の前後 5%以内である時点の割合は他の代入法を用いた場合より高くなっている。しかし、負の方向にバイアスがあるため下側 97.5%信頼区間の被覆確率が低く両方の片側 97.5%信頼区間の被覆確率が名目の信頼水準の前後 2.5%以内である時点の割合は低くなっている。右端代入法を用いた場合、両側 95%信頼区間の被覆確率が名目の信頼水準の前後 5%以内である時点の割合は中点代入法を用いた場合より低くなっているが、両方の片側 97.5%信頼区間の被覆確率が名目の信頼水準の前後 2.5%以内である時点の割合は中点代入法を用いた場合より高くなっている。しかし、最終検査時点以降は負の方向にバイアスが大きくなるため下側 97.5%信頼区間の被覆確率は低くなっている。

パラメータの組み合わせによっては大部分の片側 97.5%信頼区間の被覆確率が名目の信頼水準に近い場合があった。これらのケースを表 13 にまとめた。左端代入法を用いて区間推定する場合、被験者数が 25 人以上であれば 1 回目の検査時点から最終検査時点までの上側 97.5%信頼区間は信用できるが、下側 97.5%信頼区間は名目通りの信頼水準をもつ信頼区間より狭くなることに注意する必要がある。

中点代入法を用いて区間推定する場合、被験者数が 50 人以上であるか被験者数が 25

人で形状パラメータが低いのであれば、6 週から最終検査時点までの上側 97.5%信頼区間は信用できる。また、被験者数が 25 人以上で右側打ち切りの割合が 0.13 の場合、各検査時点から数週後までの下側 97.5%信頼区間は信用できる。しかし、右側打ち切りの割合や形状パラメータが高い場合は下側 97.5%信頼区間が名目通りの信頼水準をもつ信頼区間より狭くなることに注意が必要である。

右端代入法を用いて区間推定する場合は、被験者数が 25 人の場合で形状パラメータが低いときは 1 回目の検査時点以降、形状パラメータが 1.5 のときは 2 回目の検査時点以降の上側 97.5%信頼区間は信用できる。また、被験者数が 50 人の場合、右側打ち切りの割合が 0.5 ならば 1 回目の検査時点以降の上側 97.5%信頼区間は信用できる。右側打ち切りの割合が低いときは、ベースラインから最終検査時点までの下側 97.5%信頼区間は信用できる。以上のことから、右端代入法において、両方の片側 97.5%信頼区間が信用できる場合として、以下の場合を挙げることができる。

- ① 被験者数が 25 人で右側打ち切りの割合が低くかつ形状パラメータが低いときは、1 回目の検査時点から信用できる。
- ② 被験者数が 25 人で右側打ち切りの割合が低くかつ形状パラメータが 1.5 のときは、2 回目の検査時点から最終検査時点まで信用できる。

しかし、最終検査時点以降の下側 97.5%信頼区間やベースラインから 1 回目の検査時点までの上側 97.5%信頼区間は、名目の信頼水準をもつ信頼区間より狭く構成していることに注意が必要である。また、上記の区間以外やパラメータの組み合わせによっては被覆確率が大幅に低いこともあり、名目通りの信頼水準をもつ信頼区間より狭くなる可能性があることに注意が必要である。

例えば、上側 97.5%信頼区間の被覆確率が名目の信頼水準に近く、下側 97.5%信頼区間の被覆確率が名目の信頼水準より低い時点で区間推定をする場合、構成した信頼区間の下端は信用し、名目通りの信頼水準をもつ信頼区間の上端は構成した信頼区間よりも上にあると解釈するのがよい。

表 13 片側 97.5%信頼区間の被覆確率が名目の信頼水準に近い時点とパラメータの組み合わせ

代入法	被覆確率	条件	時点
左端代入法	上側97.5%信頼区間の被覆確率	25人以上	1回目の検査時点から最終検査時点まで
	下側97.5%信頼区間の被覆確率	該当なし	
中点代入法	上側97.5%信頼区間の被覆確率	50人以上	6週から最終検査時点まで
		25人以上 形状パラメータが低い	6週から最終検査時点まで
	下側97.5%信頼区間の被覆確率	25人以上 右側打ち切りの割合が0.13	各検査時点から数週後まで
右端代入法	上側97.5%信頼区間の被覆確率	25人	形状パラメータが低い 1回目の検査時点以降
			形状パラメータが1.5 2回目の検査時点以降
		50人 右側打ち切りの割合が0.5	1回目の検査時点以降
	下側97.5%信頼区間の被覆確率	右側打ち切りの割合が低い	ベースラインから最終検査時点まで

表 9 から最終検査時点 (60 週) のみを見ると、ほとんどのパラメータの組み合わせで上側 97.5%信頼区間の被覆確率が名目の信頼水準に近いことが分かる。また、他の代入法より左端代入法を用いた場合の被覆確率が名目の信頼水準に近いパラメータの組み合わせが多いことが分かる。また、左端代入法を用いた場合は被験者数が 50 人または 100 人の場合、どのパラメータの組み合わせでも両側、片側信頼区間の被覆確率とともに名目の信頼水準に近かった。特に被験者数が 100 人の場合、ほとんどが両側 95%信頼区間の被覆確率が名目の信頼水準の前後 1%以内かつ両方の片側 97.5%信頼区間の被覆確率が名目の信頼水準の前後 0.5%以内であった。しかし、本研究では最終検査時点での無増悪生存率が 40%になるようにパラメータを設定している。そのため、最終検査時点での無増悪生存率がさらに低くなる場合、生存率の最低値が 0 であるために分布が偏り、下側 97.5%信頼区間の被覆確率が低くなる可能性がある。

本研究のシミュレーションは実際の臨床試験を想定しており、本研究で得られた結果は、時間のスケールを変えることで様々な臨床試験の解釈に用いることができると考えられる。40%の時点が 5 回目の検査 (それまで等間隔の検査) になる臨床試験の場合は、シミュレーションで設定した右側打ち切りが起こる場合は、シミュレーションの結果をそのまま当てはめることができる。

そうでない検査タイミングの場合、例えば、図 1 の臨床試験における chemotherapy の生存率曲線では、40%点の時点はおおよそ 10~12 週であり、その時点までに検査は 6 週おきにベースラインを含め 3 回行われている。これは、本研究のシミュレーションにおける検査間隔を約 2 倍にしたものに相当する。従って、もし 40%時点までに同様の右側打ち切りが生じていれば、この臨床試験の被覆確率は、本研究のシミュレーションにおける被覆確率よりも低くなるものと推定できる。

また、図 2 の臨床試験では 3 年が経過するまで 6 か月ごとに検査し、その後は 1 年ごとに検査している。図 2 の生存率曲線から、10 年経過時の無増悪生存率がおおよそ 80%であることから、無増悪生存率が 40%点になるには、おおよそ 30 年かかると考えられる。従って、この 30 年の間の検査回数は $3 \cdot 2 + 27 = 33$ 回となり、本研究のシミュレーションの設定である 5 回の検査回数よりも非常に多く、本研究の結果より被覆確率は高くなるものと考えられる。

5 結論と今後の課題

GWSE は被験者数が多い場合、右端代入法を用いた場合の最終検査時点以降を除いて SD を正しく推定できていた。しかし、SD を正しく推定できていても被覆確率が名目の信頼水準に近かったのは一部のみであった。パラメータの組み合わせによっては被覆確率が名目の信頼水準を大きく下回ることもある。臨床試験では右端代入法がよく使われるが、最終検査時点のみで区間推定するのであれば、左端代入法を用いて感度分析を行うのが良い。他の時点で区間推定する場合は、信頼区間の上側下側ごとに表 13 に該当する時点では、信頼区間の片側は信用して良いと考えられる。一方、表 13 に該当しない場合は、信頼区間の片側は名目の信頼区間よりも広くなると解釈することが望ましい。

本研究では最終検査時点の無増悪生存率を 40%と固定している。最終検査時点の無増悪生存率がさらに低い場合、被覆確率が低くなる可能性が考えられる。

今後の課題としては、今回のシミュレーションで仮定した最終検査時点での無増悪生存率や死亡の割合、検査間隔等を変化させて推定精度を評価すること、及び他の手法を用いた場合の区間推定の性能を評価することなどが考えられる。

6 参考文献

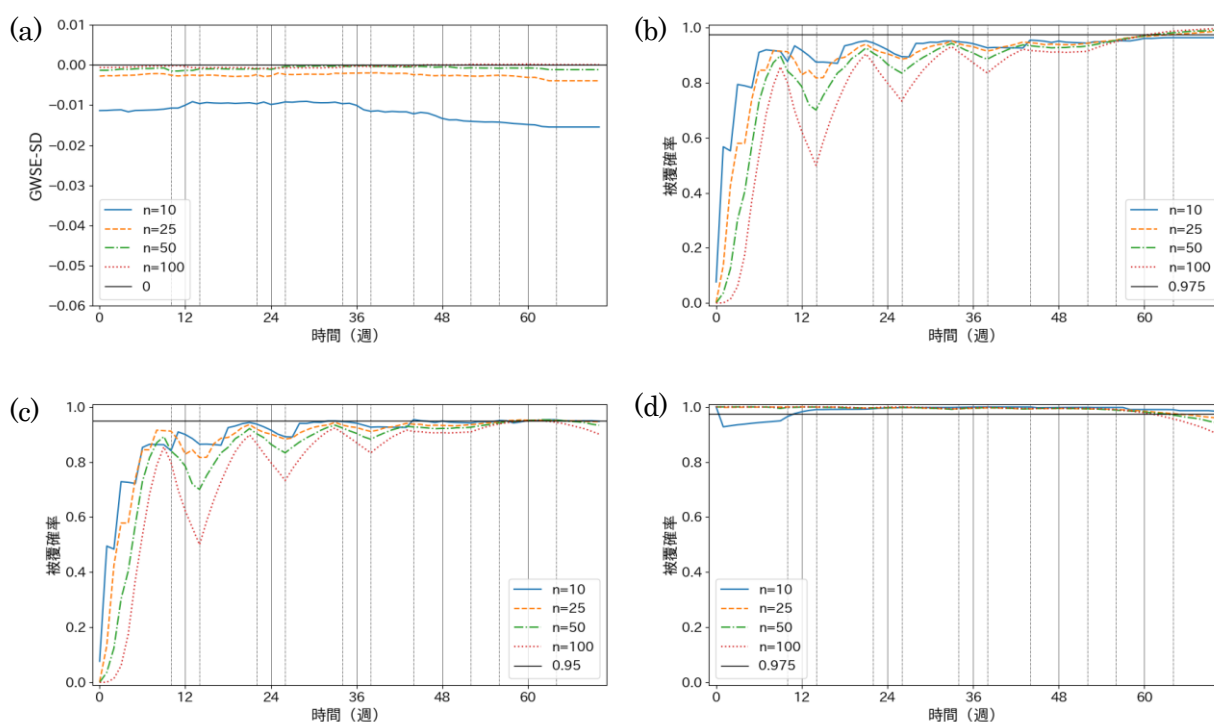
- [1] 西川正子. (2019). カプラン・マイヤー法：生存時間解析の基本手法. 共立出版, (統計学 One Point).
- [2] Law CG, Brookmeyer R. *Statistics in Medicine* 11:1569-78, 1992.
- [3] Zhang et al.(2009).*Journal of Statistical Computation and Simulation* 79(10):1245-1257.
- [4] Gauvreau K, Degruittola V, Pagano M, Bellocco R. *Statistics in Medicine* 13:2021-30, 1994.
- [5] Rubin, D. B. (1987). *Multiple Imputation for nonresponse in surveys*. John Wiley.
- [6] Taylor et al. (1990). *Statistics in Medicine* 9; 505-514, 1990.
- [7] Pan W. *Statistics in Medicine* 19:1-11, 2000.

- [8] Sun, J. (2005). *The Statistical Analysis of Interval-censored Failure Time Data*. Springer.
- [9] Ding-Geng Chen, S. Jianguo and K. E. Peace. ed (2013). *Interval-censored Time-to-Event Data*. CRC Press.
- [10] 西川正子. 区間打ち切りデータの解析. In: 医学統計学の事典(丹後俊郎, 小西貞則, 編集). 東京：朝倉書店； p.252-255. 2010.
- [11] 西川正子. 区間打ち切りデータの解析. In:新版 医学統計学ハンドブック (丹後俊郎, 松井茂之編集). 東京：朝倉書店； p.207-210. 2018.
- [12] Flaherty, K. T. et al. (2012). Improve survival with MEK inhibition in BRAF-Mutated melanoma. *The New England Journal of Medicine*, Vol. 367, 107-114.
- [13] Giuliano, A. E. et al. (2017). Effect of axillary dissection vs no axillary dissection on 10-year overall survival among women with invasive breast cancer and sentinel node metastasis: The ACOSOG Z0011 (Alliance) Randomized Clinical Trial. *JAMA*, Vol. 318, 918-926.
- [14] Nishikawa, M. and Tango, T. (2003a). Behavior of the Kaplan-Meier estimator for deterministic imputations to interval-censored data and the Turnbull estimator. *Japanese Journal of Biometrics*, Vol. 24, 71-94.
- [15] Nishikawa, M. and Tango, T. (2003b). *Statistics and Probability Letters* 65, 353-361.
- [16] Uno et al. (2013). Moving Beyond the Hazard Ratio in Quantifying the Between-Group Difference in Survival Analysis. *J Clin Oncol* 32:2380-2385.

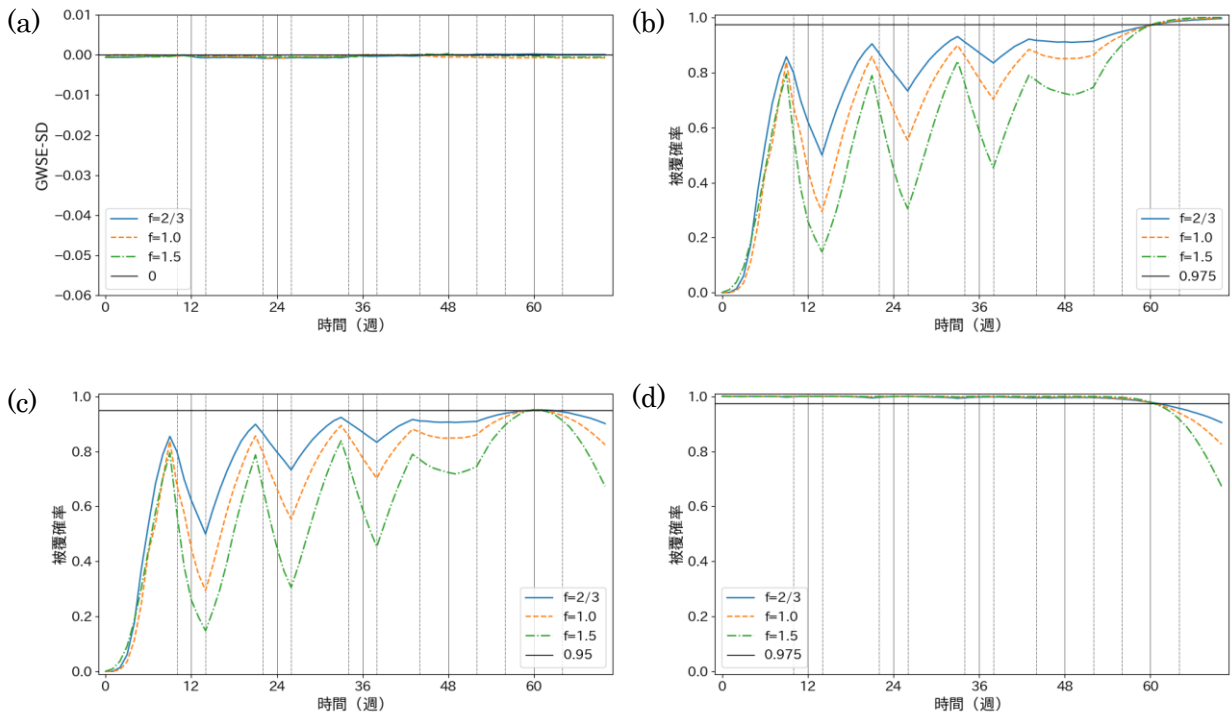
- [17] Bo Huang, Pei - Fen Kuan. (2017). Comparison of the restricted mean survival time with the hazard ratio in superiority trials with a time - to - event end point. *Pharmaceutical Statistics* 17, 3 <https://doi.org/10.1002/pst.1846>
- [18] Sarosdy, M. F., Schellhammer, P. F., Sharifi, R., Block, N. L., Soloway, M. S., Venner, P. M., Patterson, A. L., Vogelzang, N. J., Chodak, G. W., Klein, E. A., Schellenger, J. J. and Kolvenbag, G. C. M. (1998). Comparison of Goserelin and Leuprolide in combined androgen blockade therapy. *Urology* 52, 1, 82-88.

7 付録

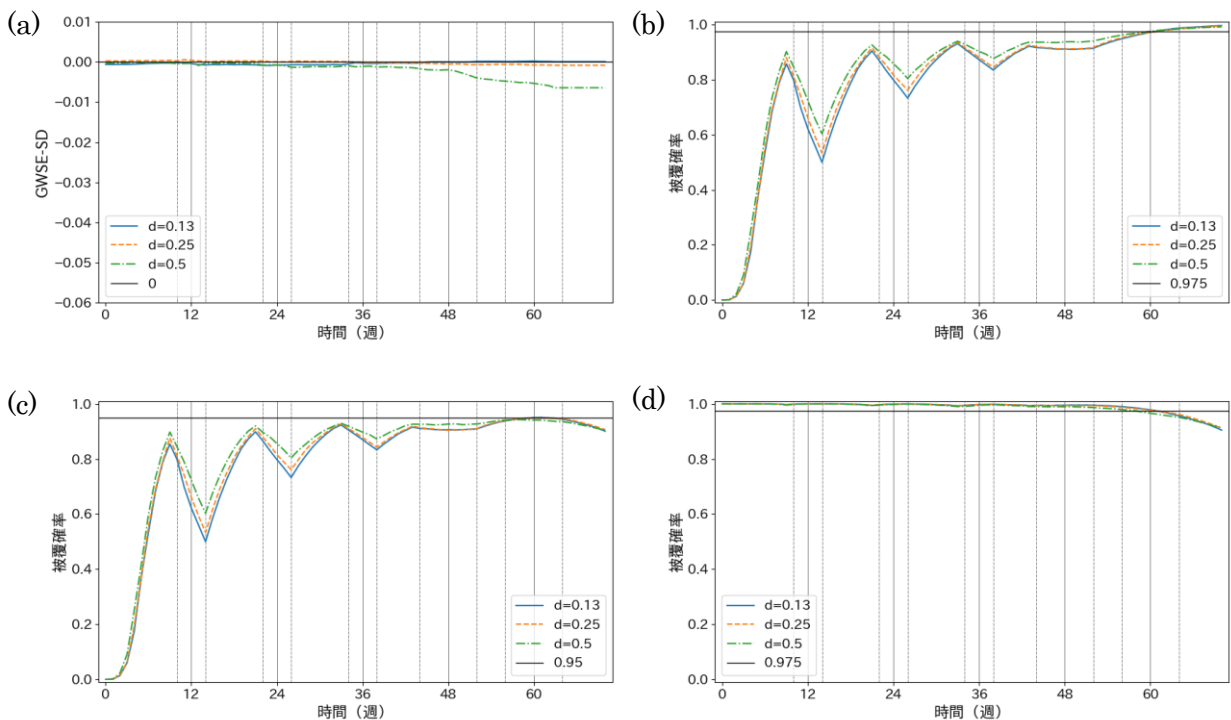
GWSE と SD の差及び被覆確率の比較結果を、様々なパラメータ値において付録図 1～付録図 6 に示した。被験者数ごとの、形状パラメータごとの、及び右側打ち切りの割合ごとの「GWSE と SD の差及び被覆確率の比較」について、左端代入法の場合を付録図 1～付録図 3 に示した。また、右端代入法の場合の同様のグラフを、付録図 4～付録図 6 に示した。



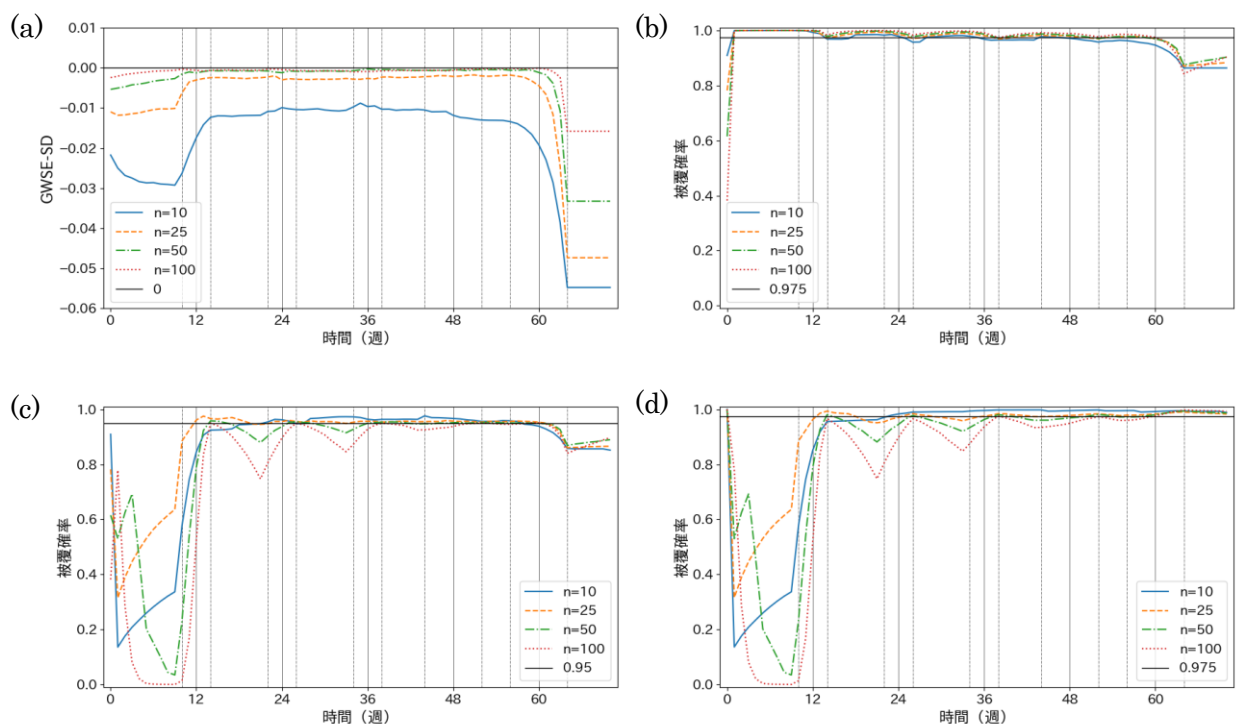
付録図 1 被験者数ごとの GWSE と SD の差及び被覆確率の比較 (左端代入法、形状パラメータ = 2/3、右側打ち切りの割合 = 0.13、 n : 被験者数、(a) GWSE と SD の差、(b) 下側 97.5%信頼区間の被覆確率、(c) 両側 95%信頼区間の被覆確率、(d) 上側 97.5%信頼区間の被覆確率)



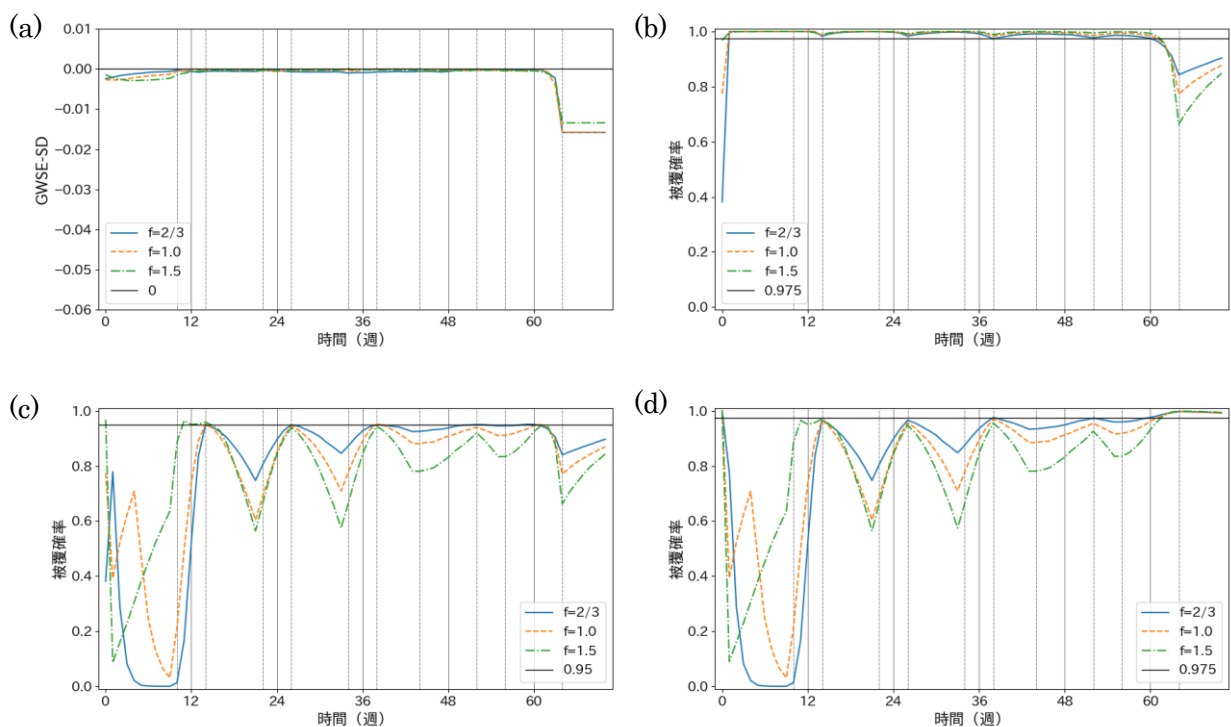
付録図 2 形状パラメータごとの GWSE と SD の差及び被覆確率の比較 (左端代入法、被験者数＝100、右側打ち切りの割合＝0.13、 f : 形状パラメータ、(a) GWSE と SD の差、(b) 下側 97.5%信頼区間の被覆確率、(c) 両側 95%信頼区間の被覆確率、(d) 上側 97.5%信頼区間の被覆確率)



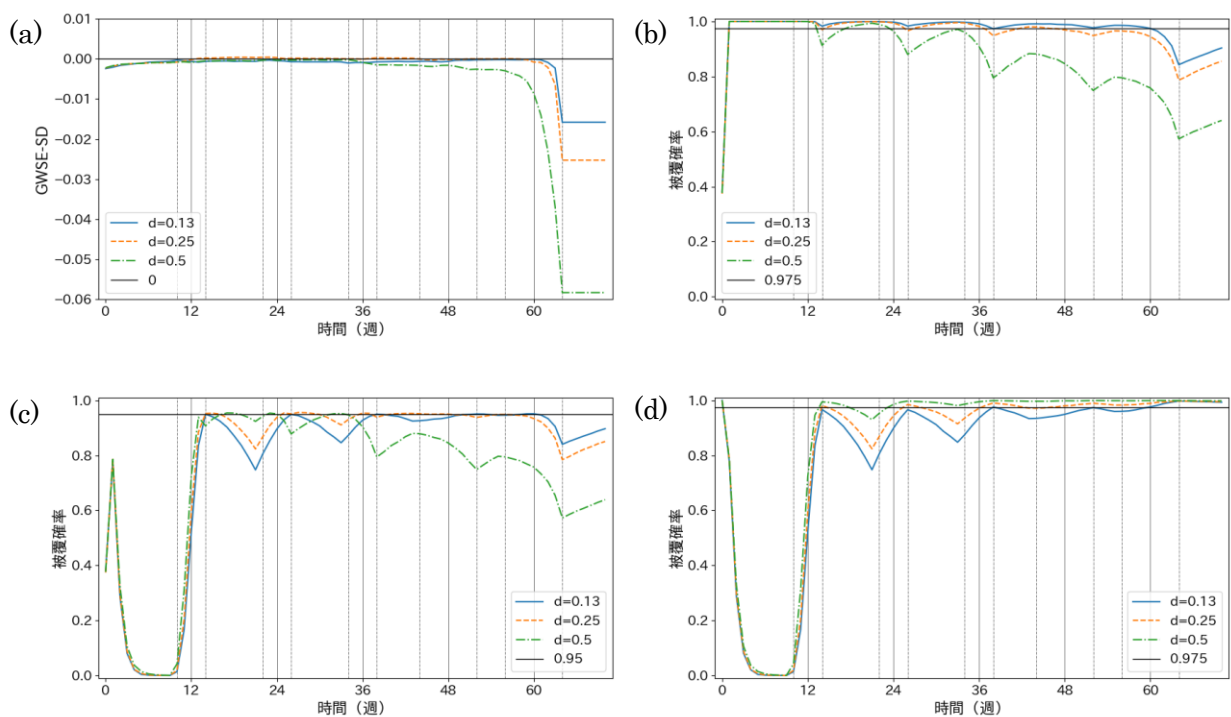
付録図 3 右側打ち切りの割合ごとの GWSE と SD の差及び被覆確率の比較 (左端代入法、被験者数＝100、形状パラメータ＝2/3、 d : 右側打ち切りの割合、(a) GWSE と SD の差、(b) 下側 97.5%信頼区間の被覆確率、(c) 両側 95%信頼区間の被覆確率、(d) 上側 97.5%信頼区間の被覆確率)



付録図 4 被験者数ごとの GWSE と SD の差及び被覆確率の比較 (右端代入法、形状パラメータ = $2/3$ 、右側打ち切りの割合 = 0.13、 n : 被験者数、(a) GWSE と SD の差、(b) 下側 97.5%信頼区間の被覆確率、(c) 両側 95%信頼区間の被覆確率、(d) 上側 97.5%信頼区間の被覆確率)



付録図 5 形状パラメータごとの GWSE と SD の差及び被覆確率の比較 (右端代入法、被験者数 = 100、右側打ち切りの割合 = 0.13、 f : 形状パラメータ、(a) GWSE と SD の差、(b) 下側 97.5%信頼区間の被覆確率、(c) 両側 95%信頼区間の被覆確率、(d) 上側 97.5%信頼区間の被覆確率)



付録図 6 右側打ち切りの割合ごとの GWSE と SD の差及び被覆確率の比較（右端代入法、被験者数=100、形状パラメータ=2/3、 d : 右側打ち切りの割合、(a) GWSE と SD の差、(b) 下側 97.5% 信頼区間の被覆確率、(c) 両側 95%信頼区間の被覆確率、(d) 上側 97.5%信頼区間の被覆確率)

(原稿提出: 2019 年 1 月 20 日; 修正稿提出: 2020 年 2 月 9 日)